
WPIREH

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligent Genetics, Inc.

MPSrch_n n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sat Jun 28 13:09:59 1997; MasPar time 2747.71 Seconds
Tabular output not generated. 1214.018 Million cell updates/sec

Title: >US-08-731-499-1
Description: (1-3000) from US08731499.seq
Perfect Score: 3000
N.A. Sequence: 1 CGCGCGCGCGCGCGCTGGAACAAGAAAAA 3000
Comp: GCGCGCGCGCGCGCGGACC.....TTTGTCTTTTTTTTTT

Scoring table: TABLE default
Gap 6
Nmatch STD : Dbase 0; Query 0
Searched: 333249 seqs, 555961234 bases x 2
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-new11
1: BCT 2: FOUN 3: GEN 4: HUM1 5: HUM2 6: HUM3 7: INV1 8: INV2
9: INV3 10: INV4 11: INV5 12: INV6 13: INV7 14: INV8 15: INV9
16: VRT 17: PLN 18: PRO1 19: PRO2 20: ROD 21: SYN 22: UNC
23: VIR1 24: VIR2
genbank97

Database: 25: BCT1 26: BCT2 27: BCT3 28: BCT4 29: BCT5 30: BCT6 31: BCT7
32: BCT8 33: BCT9 34: GEN1 35: GEN2 36: HTG 37: INV1 38: INV2
39: INV3 40: INV4 41: INV5 42: INV6 43: INV7 44: INV8 45: INV9
46: MAM1 47: MAM2 48: MAM3 49: VRT1 50: VRT2 51: VRT3 52: PAT1
53: PAT2 54: PAT3 55: PAT4 56: PHG 57: PLN1 58: PLN2 59: PLN3
60: PLN4 61: PLN5 62: PLN6 63: PLN7 64: PLN8 65: PLN9 66: PLN10
67: PR11 68: PR12 69: PR13 70: PR14 71: PR15 72: PR16 73: PR17
74: PR18 75: PR19 76: PR110 77: PR111 78: PR112 79: PR113
80: PR114 81: ROD1 82: ROD2 83: ROD3 84: ROD4 85: ROD5 86: ROD6
87: ROD7 88: ROD8 89: STR 90: SYN 91: UNA 92: VRL1 93: VRL2
94: VRL3 95: VRL4 96: VRL5 97: VRL6 98: VRL7 99: VRL8 100: VRL9
genbank-new11
101: BCT 102: GEN 103: INV1 104: INV2 105: MAM 106: VRT
107: PHG 108: PLN 109: PR11 110: ROD 111: ROD 112: SYN
113: UNA 114: VRL
u-emb148.97
115: part1 116: part2

Database: 101: BCT 102: GEN 103: INV1 104: INV2 105: MAM 106: VRT
107: PHG 108: PLN 109: PR11 110: ROD 111: ROD 112: SYN
113: UNA 114: VRL
u-emb148.97
115: part1 116: part2

Statistics: Mean 12.787; Variance 8.344; scale 1.532
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result Query
No. Score Match Length DB ID Description Pred. No.
SUMMARIES

	1	3000	100.0	3000	73	HSU02680	Human protein tyrosin	0.00e+00
	2	35	1.2	356	61	PSNOD6MR	P.sativum PSNOD6 MRNA	2.01e-02
	3	35	1.2	1338	47	ECCD44	E.caballus MRNA for l	2.01e-02
C	4	36	1.2	1508	44	S5234	CAR2-CAMP receptor su	7.24e-03
	5	35	1.2	1511	60	LEH1HSTL	L.esculentum MRNA for	2.01e-02
	6	36	1.2	1891	103	DDU73685	Dictyostelium discoid	7.24e-03
	7	37	1.2	2130	85	MUSP53PG	Mouse p53 cellular tu	2.57e-03
	8	37	1.2	2130	20	MMKP53PG	Mouse p53 cellular tu	2.57e-03
	9	37	1.2	2132	82	MMP53P	Mouse pseudogene for	2.57e-03
	10	36	1.2	2834	72	HSPMSCL	H.sapiens MRNA for PM	7.24e-03
	11	35	1.2	3402	44	S73909	abMa-actin-based moto	2.01e-02
	12	32	1.1	350	55	I25442	Sequence 11 from pate	2.95e-01
C	13	34	1.1	542	44	PNRSUB	P.falciiparum non-repe	5.51e-02
	14	33	1.1	795	86	MUSUNKNL	Mouse (clone HORT13)	1.49e-01
C	15	33	1.1	840	59	DDIHPERB	D.discoidium protein	1.49e-01
	16	32	1.1	1225	13	DDU66910	Dictyostelium discoid	3.95e-01
	17	34	1.1	1246	57	ATPRXR4GE	A.thaliana MRNA for p	5.51e-02
	18	34	1.1	1247	57	ATPOL	A.thaliana MRNA for p	5.51e-02
	19	34	1.1	1260	61	PPPHRG	P.polycephalum PHP ge	5.51e-02
	20	34	1.1	1260	115	S41278	php-cell type-specific	5.51e-02
	21	32	1.1	1444	43	HELMAMIDE	H.echinata MRNA for L	3.95e-01
	22	33	1.1	1540	54	I08701	Sequence 4 from Paten	1.49e-01
	23	32	1.1	1549	23	FQJRNAP	foot and mouth diseas	3.95e-01
	24	32	1.1	1549	92	APHRNAP	foot and mouth diseas	3.95e-01
	25	33	1.1	1563	40	DDPYRL3A	D.discoidium PYRL-3A	1.49e-01
	26	34	1.1	1606	13	DDU41222	Dictyostelium discoid	5.51e-02
	27	34	1.1	1606	40	DDU41222	Dictyostelium discoid	5.51e-02
	28	32	1.1	1645	60	LEVSF1	L.esculentum vsf-1 MR	3.95e-01
	29	34	1.1	1659	40	CVCAP	C.viridissima cap MRN	5.51e-02
	30	33	1.1	1864	115	ATLBAAP3	A.thaliana Landbergis	1.49e-01
C	31	33	1.1	1898	40	DDDDGI	Dictyostelium discoid	1.49e-01
	32	34	1.1	1908	82	MMRNASEB4	M.musculus seb4 MRNA	5.51e-02
	33	34	1.1	1990	53	AI7010	tomato fruit pectin e	5.51e-02
C	34	33	1.1	2033	59	DDI1091B	Dictyostelium discoid	1.49e-01
C	35	33	1.1	2383	44	PFALD	P.falciiparum aldolase	1.49e-01
C	36	33	1.1	2383	52	AI3481	P.falciiparum gene for	1.49e-01
	37	32	1.1	2647	64	SOPKAA	S.oleracea MRNA for p	3.95e-01
	38	32	1.1	2794	40	DDU48271	Dictyostelium discoid	3.95e-01
	39	33	1.1	2831	65	STU60202	Solanum tuberosum lip	1.49e-01
	40	33	1.1	2865	76	HUMERF72H	Human (clone PA3) pro	1.49e-01
	41	33	1.1	2982	59	DDISGSPA	Dictyostelium discoid	1.49e-01
	42	33	1.1	3437	64	SOPULSPO	S.oleracea L. MRNA fo	1.49e-01
	43	32	1.1	3548	73	HSSPR2	H.sapiens SPR-2 MRNA	3.95e-01
	44	33	1.1	3792	83	MMU63323	Mus musculus translat	1.49e-01
	45	33	1.1	5001	44	PFAABRA	Plasmodium falciparum	1.49e-01

ALIGNMENTS

RESULT 1 HSU02680 3000 bp MRNA PRI 03-FEB-1994
LOCUS Human protein tyrosine kinase mRNA, complete cds.
DEFINITION
ACCESSION U02680
NID 9451481

SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 3000)
Beeler, J.F., Larochelle, W.J., Chedid, M., Tronick, S.R. and
Aaronson, S.A.

TITLE Prokaryotic expression cloning of a novel human tyrosine kinase
JOURNAL Mol. Cell. Biol. 14 (2), 982-988 (1994)
MEDLINE 94119116
REFERENCE 2 (bases 1 to 3000)
Beeler, J.F.
AUTHORS Direct Submission
TITLE Submitted (22-OCT-1993) John F. Beeler, Laboratory of Cellular &
JOURNAL Molecular Biology, National Cancer Institute, Building 37 Room
1E24, Bethesda, MD 20892, USA
FEATURES Location/Qualifiers

QY 2953 ATATAAAATTTTCTCTTTAAAAAAGAAAACAGAAGAAAAA 3000

7
RESULT
LOCUS MUSEP53PG 2130 bp DNA ROD 03-MAY-1985
DEFINITION Mouse p53 cellular tumour antigen psuedogene.
ACCESSION K02110
NID 9200206
KEYWORDS antigen; p53 gene; processed pseudogene; tumor antigen.
SOURCE Mouse 3.3-kb fragment isolated from a BALB/c genomic library, clone pCh53-11.

ORGANISM
Mus musculus
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Mus.

REFERENCE
1 (bases 1 to 2130)
Zakut-Houri, R., Oren, M., Bienz, B., Lavie, V., Hazum, S. and Givol, D.
A single gene and a pseudogene for the cellular tumour antigen p53
Nature 306, 594-597 (1983)

JOURNAL
MEDLINE
COMMENT
The DNA sequence of pCh53-11 contains a long poly-A tract, lacks introns, and is bounded by direct repeats (bp 169-181 and bp 1852-1864), suggesting that it is a processed gene which resulted from reverse transcription of the mature mRNA.
The sequence of the murine p53 cDNA, also reported by [1] (see separate entry), and the pseudogene are almost identical from nucleotide 186 onward. Upstream of this position the two sequences diverge totally and no homology can be observed; downstream of here the two sequences differ by only 4%. The differences are due to substitutions and to some small deletions or additions in this gene relative to the cDNA.

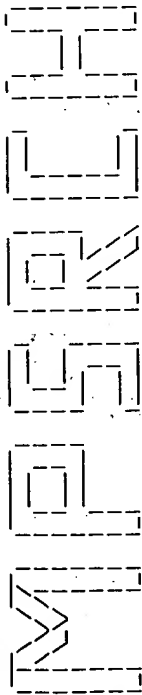
FEATURES
Location/Qualifiers
1..2130
/organism="Mus musculus"
<1..1826
/note="pseudo-p53 mRNA"
262..>1434
/note="pseudo-p53"
/pseudo
/codon_start=1
602 a 585 c 487 t
BASE COUNT 19 bp upstream of BglII site.

ORIGIN
Query Match 1.2% Score 37; DB 85; Length 2130;
Best Local Similarity 82.5% Pred. No. 2.57e-03;
Matches 47; Conservatve 0; Mismatches 10; Indels 0; Gaps 0;

Db 1798 ttatttcacataaaatttgccttatcactaaaaaagaaaaaaaaaaaaa 1854
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2944 TTATATCACATATAAAATTTTTTCTCTTTAAAAAAGAAAACAGAAGAAAAA 3000

RESULT 8
ID MKP53PG standard; DNA; ROD; 2130 BP.
AC K02110;
NCI 9200206
NI 01-OCT-1996 (Rel. 49, Created)
DT DT 01-OCT-1996 (Rel. 49, Last updated, Version 1)
DE Mouse p53 cellular tumour antigen psuedogene.
KW antigen; p53 gene; processed pseudogene; tumor antigen.
OS Mus musculus (mouse)
CC Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
OC Theria; Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
RN [1]
RP 1-2130
RX MEDLINE; 84068204.
RA Zakut-Houri R., Oren M., Bienz B., Lavie V., Hazum S., Givol D.;
RT "A single gene and a pseudogene for the cellular tumour antigen p53";
RL Nature 306:594-597(1983).
CC The DNA sequence of pch53-11 contains a long poly-A tract, lacks introns, and is bounded by direct repeats (bp 169-181 and bp 1852-1864), suggesting that it is a processed gene which resulted

8/25/97



(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligent, Inc.
MPSrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sat Jun 28 19:53:43 1997; MasPar time 297.43 Seconds
Tabular output not generated. 877.421 Million cell updates/sec
Title: >US-08-731-499-6
Description: (1-2821) from US08731499.seq
Perfect Score: 2821
N.A. Sequence: 1 ATCCCTAAGACGACACGCTG.....AGGGTTCAAGACACGCTG 2821
Comp: TAGGATCTGCGTGGGAC.....TCCCCAAGTTCGTCGGAC

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0
Searched: 121476 seqs, 46255616 bases x 2
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq26
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23

Statistics: Mean 10.232; Variance 6.928; scale 1.477

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	234	8.3	8342	13	Q75209	ALL-1 (acute lymphoc	1.18e-120
2	234	8.3	22481	23	T11658	PEDF full length sequ	1.18e-120
3	228	8.1	1496	1	N92386	DNA coding for the si	5.78e-117
4	224	7.9	1470	7	Q47355	Myotonic dystrophy ge	1.66e-114
5	221	7.8	1494	1	Q03369	Sequence encoding hum	1.15e-112
6	217	7.7	8391	18	T16333	MLL gene 8.3 kb fragm	3.27e-110
7	216	7.7	15328	13	O81139	HPLA2-8 gene.	1.34e-109
8	218	7.7	22481	23	T11658	PEDF full length sequ	7.97e-111
9	213	7.6	1368	23	T33580	MHC region between HL	9.26e-108
10	215	7.6	8392	9	Q53478	MLL gene 8.3 kb BamHI	5.51e-109
11	208	7.4	6063	6	Q37205	Delta-amino leucinat	1.07e-104
12	207	7.3	2581	2	N70974	Sequence of human atr	4.37e-104
13	207	7.3	5836	15	Q96296	Human prostrate-speci	4.37e-104
14	207	7.3	17327	7	Q44278	Serglycin - proteogly	4.37e-104
15	204	7.2	1721	23	T33752	Macaque mucosal addre	2.99e-102
16	201	7.1	1688	11	Q62613	Human mdr-1 promoter	2.04e-100
17	201	7.1	2090	11	Q62612	Human mdr-1 promoter	2.04e-100

C 18	199	7.1	3647	9	Q56909	DNA encoding a glycos	3.40e-99
C 19	201	7.1	6404	7	Q42388	6.5 kb BamHI fragmant	2.04e-100
C 20	199	7.1	19012	6	Q36952	HSA gene.	3.40e-99
C 21	197	7.0	3647	2	Q13333	GDP-Fuc:beta-D-Gal(1,	5.65e-98
C 22	195	6.9	2649	6	Q35034	DNA fragment contg. A	9.38e-97
C 23	196	6.9	10897	19	T09187	Mutu putative oncogen	2.30e-97
C 24	191	6.8	2589	7	Q41454	Human pre-proANVP ge	2.58e-94
C 25	191	6.8	2590	3	N50695	Sequence encoding pre	2.58e-94
C 26	193	6.8	11357	9	Q51024	Human FEER1 beta chai	1.56e-95
C 27	188	6.7	3096	2	N71063	Genomic sequence enco	1.73e-92
C 28	188	6.7	8174	2	Q13332	GDP-Fuc:beta-D-galact	1.73e-92
C 29	188	6.7	8174	2	Q56908	DNA encoding a glycos	1.73e-92
C 30	190	6.7	13747	7	Q47342	Myotonic dystrophy ge	1.05e-93
C 31	187	6.6	361	8	Q60655	Human brain Expressed	7.04e-92
C 32	185	6.6	1759	3	Q23030	Clone W211 encoding c	1.16e-90
C 33	187	6.6	10897	19	T09187	Mutu putative oncogen	7.04e-92
C 34	187	6.6	17350	12	Q72708	hOP1 human osteogenic	7.04e-92
C 35	187	6.6	17410	9	T18381	hOP-1 genomic DNA.	7.04e-92
C 36	187	6.6	17410	9	Q53142	Sequence encoding ost	7.04e-92
C 37	187	6.6	17415	17	T11639	Human osteogenic prot	7.04e-92
C 38	182	6.5	328	8	Q60486	Human brain Expressed	7.75e-89
C 39	184	6.5	352	8	Q60764	Human brain Expressed	4.71e-90
C 40	183	6.5	426	8	Q59859	Human brain Expressed	1.91e-89
C 41	183	6.5	3480	5	Q28568	Human thrombin recept	1.91e-89
C 42	184	6.5	5108	2	N60388	Sequence encoding hum	4.71e-90
C 43	184	6.5	5468	1	Q06340	Human low density lip	4.71e-90
C 44	184	6.5	7620	6	Q39286	Glucocerebrosidase ge	4.71e-90
C 45	181	6.4	30967	23	T32454	Calpain large subunit	3.14e-88

ALIGNMENTS

RESULT	1	
ID	Q75209	standard; cDNA; 8342 BP.
AC	Q75209;	
DE	23-AUG-1995	(first entry)
DE	ALL-1	(acute lymphocytic leukaemia-1) breakpoint cluster region.
KW	Acute lymphoblastic leukaemia;	acute nonlymphoblastic leukaemia;
KW	Chromosomal translocation;	abnormality; detection; rearrangement;
KW	breakpoint cluster region;	Alu repeat; chromosome 11; probe B859; ds.
OS	Hom sapiens.	
PH	Key	Location/Qualifiers
FT	exon	1..263
FT	/tag= a	
FT	/number= 5	
FT	Intron	
FT	/tag= b	264..2352
FT	exon	
FT	/tag= c	593..666
FT	/number= 6	
FT	Intron	
FT	/tag= d	667..798
FT	repeat_unit	
FT	/tag= e	799..1108
FT	/rpt_type= OTHER	
FT	/note= "Alu repeat-a (Class J)"	
FT	repeat_unit	1119..1420
FT	/tag= f	
FT	/rpt_type= OTHER	
FT	/note= "Alu repeat-b (Class Sx)"	
FT	repeat_unit	1432..1716
FT	/tag= g	
FT	/rpt_type= OTHER	
FT	/note= "Alu repeat-c (Class Sb0)"	
FT	repeat_unit	1921..2216
FT	/tag= h	
FT	/rpt_type= OTHER	
FT	/note= "Alu repeat-d (Class J)"	
FT	exon	2353..2484
FT	/tag= i	
FT	/number= 7	
FT	Intron	
FT	/tag= j	2485..3031

6/20/18

p. 1597 CTCTGATCTGCCCGCCCTCGGCCCTCCCAAGTGCCTGGGATTACAGGCGTGAGCCACCCGTTG 1538

— — — — —

[illegible]

MPSRELH

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Jun 28 21:17:47 1997; MasPar time 131.85 Seconds
Tabular output not generated. 845.467 Million cell updates/sec

Title: >US-08-731-499-7
Description: (1-1205) from US08731499.seq
Perfect Score: 1203
M.A. Sequence: 1 GCAGCGGTGAGTCCGCCGCC.....TCAGTGTGTTGGTTAAGTGG 1205
M.A. Comp: CGCGCGCACTCAGCGCGGGG.....AGTCACAAACCAATTCACC

Scoring table: TABLE default
Gap 6

Match - STD : Dbase 0; Query 0

Searched: 121476 seqs, 46255616 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneeq26

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23

Statistics: Mean 9.351; Variance 5.459; scale 1.713

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	419	34.8	464	19	Human gene signature	2.75e-276
2	87	7.2	1047	2	Human Natriuretic Pep	1.62e-39
3	72	6.0	1047	2	Human Natriuretic Pep	1.39e-29
4	45	3.7	204	1	Base substituted E.co	1.59e-12
5	42	3.5	91	9	Oligonucleotide probe	9.77e-11
6	41	3.4	91	9	Oligonucleotide probe	3.80e-10
7	39	3.2	204	1	Base substituted E.co	5.58e-09
8	36	3.0	114	12	Generic DNA sequence	2.93e-07
9	35	2.9	114	12	Generic DNA sequence	1.07e-06
10	35	2.9	114	12	Generic DNA sequence	1.07e-06
11	35	2.9	114	12	Generic DNA sequence	1.07e-06
12	34	2.8	114	12	Generic DNA sequence	3.89e-06
13	34	2.8	114	12	Generic DNA sequence	3.89e-06
14	34	2.8	114	12	Generic DNA sequence	3.89e-06
15	34	2.8	114	12	Generic DNA sequence	3.89e-06
16	33	2.7	114	12	Generic DNA sequence	1.39e-05
17	32	2.7	114	12	Generic DNA sequence	4.91e-05

18	32	2.7	114	12	Q70466	Generic DNA sequence	4.91e-05
19	32	2.7	114	12	Q70470	Generic DNA sequence	4.91e-05
20	31	2.6	114	12	Q70472	Generic DNA sequence	1.71e-04
21	30	2.5	114	12	Q70471	Generic DNA sequence	5.87e-04
22	28	2.3	114	12	Q70473	Generic DNA sequence	6.58e-03
23	28	2.3	114	12	Q70472	Generic DNA sequence	6.58e-03
24	27	2.2	82	21	T13610	DC43 TSAR library gen	2.15e-02
25	26	2.2	111	16	T02821	MS-associated retrovi	6.88e-02
26	27	2.2	114	12	Q70471	Generic DNA sequence	2.15e-02
27	26	2.2	114	12	Q70473	Generic DNA sequence	6.88e-02
28	27	2.2	565	6	Q35072	HCV envelope region n	2.15e-02
29	25	2.1	3871	2	N71302	HSV-1 gB and surround	2.15e-01
30	24	2.0	74	21	T13613	DC43 TSAR library gen	6.61e-01
31	24	2.0	75	21	T13612	DC43 TSAR library gen	6.61e-01
32	24	2.0	81	21	T13611	DC43 TSAR library gen	6.61e-01
33	24	2.0	81	21	T13611	DC43 TSAR library gen	6.61e-01
34	24	2.0	108	22	T29064	Probe for Candida alb	6.61e-01
35	24	2.0	565	6	Q35072	HCV envelope region n	6.61e-01
36	23	1.9	39	23	T14323	Sequence used in the	1.98e+00
37	23	1.9	74	21	T13613	DC43 TSAR library gen	1.98e+00
38	23	1.9	82	21	T13610	DC43 TSAR library gen	1.98e+00
39	23	1.9	330	2	Q11188	Immunoglobulin kappa	1.98e+00
40	23	1.9	1393	17	T11061	Growth differentiatio	1.98e+00
41	23	1.9	3871	2	N71302	HSV-1 gB and surround	1.98e+00
42	23	1.9	5746	23	T09225	Partial sequence of v	1.98e+00
43	22	1.8	75	21	T13612	DC43 TSAR library gen	5.77e+00
44	22	1.8	501	3	N50028	Sequence encoding new	5.77e+00
45	22	1.8	2492	6	Q36410	HpaI restriction enzy	5.77e+00

ALIGNMENTS

RESULT 1
ID T22418 standard; cDNA to mRNA; 464 BP.

AC T22418;

DT 07-AUG-1996 (first entry)

DE Human gene signature HUMGS04025.

KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;

KW human; cloning; mapping; non-biased library; diagnosis; detection;

KW cell typing; abnormal cell function; ss.

OS Homo sapiens.

PN W09514772-A1.

PD 01-JUN-1995.

PF 11-NOV-1994; J01916.

PR 12-NOV-1993; JP-355504.

PA (MATSU) MATSUBARA K.

PI Matsubara K, Okubo K;

DR WPI; 95-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

PT tissues

PS Claim 1; Page 1120; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in T19001-T26837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

SQ Sequence 464 BP; 124 A; 74 C; 153 G; 111 T;

Query Match

34.8%; Score 419; DB 19; Length 464;

Best Local Similarity 96.7%; Pred. No. 2.75e-276;

8/20/97

W P S R L H

(TM)

Release 2.1D John F. Collins, BioComputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligent, Inc.

MPSrch_un n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Jun 28 22:40:56 1997; MasPar time 213.20 Seconds

Tabular output not generated. 867.829 Million cell updates/sec

Title: >US-08-731-499-9
Description: (1-2000) from US08731499.seq (1 of 6)
Perfect Score: 2000

N.A. Sequence: 1 CCATCAATTTCTTATTTT AATATGTTTACTAAGTGT 2000
Comp: GGTAGTATAAGATATAAAA TTTATCAATGATTTCACA

Scoring table: TABLE default.

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 121476 seqs, 46255616 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq26

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23

Statistics: Mean 9.913; Variance 8.353; scale 1.187

Ca Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	ID	Description	Pred. No.
1	135	6.8	3158	9	Q53212	Human cyclin D3 promo	1.64e-48
2	135	6.8	3158	5	Q31880	Cyclin D3 promoter.	1.64e-48
3	121	6.1	283	11	Q63862	AP2 sequence obtd. by	8.37e-42
4	120	6.0	7849	16	Q94109	hML genomic DNA.	2.51e-41
5	118	5.9	1618	7	Q46958	Human cytokine synthe	2.25e-40
6	118	5.9	1618	2	Q10207	PH15C insert containi	2.25e-40
7	118	5.9	6511	14	Q95493	Human Cdn-2 DNA.	2.25e-40
8	117	5.8	321	8	Q52028	Human brain Expressed	6.73e-40
9	117	5.8	321	6	Q39796	Expressed Sequence Ta	6.73e-40
10	116	5.8	335	8	Q60863	Human brain Expressed	2.01e-39
11	117	5.8	2600	12	N90029	Human interleukin-1 r	6.73e-40
12	117	5.8	2600	12	Q73764	Human IL-1 receptor g	6.73e-40
13	117	5.8	2600	1	N90118	CDNA of human interle	6.73e-40
14	117	5.8	2963	8	Q49333	IL-IR cDNA.	6.73e-40
15	117	5.8	3234	15	Q92781	Human thymopoietin ge	6.73e-40
16	115	5.8	11531	9	Q54222	BSSL/CEL Gene.	6.01e-39
17	117	5.8	17327	7	Q44278	Serglycin - proteogly	6.73e-40

```

18 114 5.7 4823 22 T37384 Human. thrombopoietin 1.79e-38
19 114 5.7 4823 17 T03943 Human thrombopoietin 1.79e-38
20 114 5.7 4823 16 T04051 Sequence encoding hae 1.79e-38
21 112 5.6 2435 23 T33155 Tissue plasminogen ac 1.59e-37
22 112 5.6 2888 1 Q03743 Human SKI related gen 1.59e-37
23 112 5.6 6905 15 Q92779 Human thymopoietin co 1.59e-37
24 111 5.6 8363 15 Q92408 Human cyclin A gene. 4.75e-37
25 112 5.6 17327 7 Q44278 Serglycin - proteogly 1.59e-37
26 112 5.6 2481 23 T11658 PEDF full length sequ 1.59e-37
27 110 5.5 2649 6 Q35034 DNA fragment contig. A 1.41e-36
28 110 5.5 7620 6 Q39286 Glucocerebrosidase ge 1.41e-36
29 110 5.5 10684 23 T33758 Control region isolat 1.41e-36
30 109 5.4 308 8 Q60826 Human brain Expressed 4.20e-36
31 109 5.4 743 2 N70812 Sequence encoding hum 4.20e-36
32 109 5.4 2660 3 N30022 Sequence of gene for 1.25e-35
33 108 5.4 4382 2 Q12759 P40 genomic DNA. 1.25e-35
34 108 5.4 6063 6 Q39286 Delta-amino levulinat 1.25e-35
35 108 5.4 10897 19 T09187 Mutu putative oncogen 1.25e-35
36 109 5.4 13585 17 T11549 Tumour rejection anti 4.20e-36
37 109 5.4 30967 23 T32454 Calpain large subunit 4.20e-36
38 106 5.3 366 8 Q50353 Human brain Expressed 1.10e-34
39 107 5.3 1777 12 Q79355 Sequence of the exten 3.70e-35
40 107 5.3 2320 4 Q26657 glut4 promoter/enhanc 3.70e-35
41 107 5.3 2339 2 Q10956 Encodes human 75kD TN 3.70e-35
42 107 5.3 2425 18 T11027 DNA encoding the huma 3.70e-35
43 107 5.3 2425 20 T10283 Gene for RNA componen 3.70e-35
44 106 5.3 5359 17 T12251 Cytochrome P450 isoen 1.10e-34
45 107 5.3 9272 12. Q79353 Human genomic clone h 3.70e-35

```

ALIGNMENTS

RESULT 1
ID Q53212 standard; DNA; 3158 BP.

AC Q53212;

DT 22-JUN-1994 (first entry)

DE Human cyclin D3 promoter

KW D-type; mammalian; CLN protein; protein deficiency; cell cycle start;

OS yeast; complement; ds.

KW Homo sapiens.

FT key Location/Qualifiers

FT miscfeature 3156..3158

FT /*tag= a

FT /note= "Initiation ATG codon"

FN W0324514.4.

PD 09-DEC-1993.

PF 25-MAY-1993; U05000.

PR 26-MAY-1992; US-888178.

PA (MITO-) MITOXIX.

PI Beach DH

DR WPI; 93-405720/50.

DR New D-type mammalian cyclin - replaces CLN-type protein needed

PT for cell start in budding yeast and is detected by antibodies or

PT hybridisation in biological samples to determine abnormal cell

PT division

PS Disclosure: Fig 13; 108pp; English.

CC The sequence is that of human cyclin D3 promoter.

SQ Sequence 3158 BP; 952 A; 674 C; 722 G; 810 T;

Query Match 6.8%; Score 135; DB 9; Length 3158;

Best Local Similarity 77.0%; Pred. No. 1.64e-48;

Matches 221; Conservative 0; Mismatches 62; Indels 4; Gaps 2;

Db 1653 ggcgcgggaacgtggtcaccgtgtatccacgacatttggagggccgagacggcgga 1712

Cp 293 GGACGGGTATAGTGGCTCACACCTACTACTCCATGCTTTGGGAAGCCGAGTGGTGA 234

Db 1713 tcac--gaggtcaggggttcagactacgtcggccaaacatagtgaacccactctacg 1770

Cp 233 TCACATGAGCCGAGGTGGAGACCCAGCCCTGGCCACATGATGGTAACCCCTTATCTGT 174

Db 1771 aaaaatacaaaaattagtcaggcattgggtgcgtgctgtagtcacactactcgga 1830

Cp	173	AAAAATACAAATAATGACTGGCGATGCTGGTAAATACACACTGTAATCCGACGATATTGGGAA	114
Db	1831	ttgcttgaacacggagagtgaggtgttcagtgagccgagatcgaccactgcactccagc	1890
Cp	113	TCACTTGAACCCAGAGGTGGAGGTGTCAGTCAGCCAAAGTCGACCACTGG--TCCAGC	56
Db	1891	ttgagcaacagagtgagacttcgtctcaaaaaaataaaaaaataaaaaa	1937
Cp	55	CTGGGCAACAGACAGCAAGTCTCCCTCTCCGCCCAAAAAAATAGAAA	9

RESULT 2

ID Q31880 standard; DNA; 3158 bp.

AC Q31880;

AD 22-APR-1993 (first entry)

DE Cyclin D3 promoter.

KW Cyclin; D1; D2; D3; promoter; human; liver; genomic library; clone;

KW upstream; exon; intron; neural; pCYCD1-H12; mutant; yeast; strain;

KW CLN; cyclin; gene; CLN 1; CLN 2; human; glioblastoma; cDNA library;

KW expression vector; pADNS; transformant; pCYCD1-21; pCYCD1-19; HeLa;

KW ss.

OS Homo sapiens.

PN WQ9220786-A.

PD 26-NOV-1992.

PF 18-MAY-1992; U04146.

PR 16-MAY-1991; US-701514.

PA (COLD-) COLD SPRING HARBOR LAB.

BE Beach DH;

PI WPI; 92-415774/50.

DR Recombinant mammalian D-type cyclin - replaces a CLN-type protein

PT essential for cell start in budding yeast, its antibodies and

PT probes being useful in detecting D-type cyclin in biological

PT samples

PS Disclosure; Fig 13; 75pp; English.

CC The sequences given in Q31878-80 represents the cyclin D1 to D3

CC promoters. These sequences were identified during the isolation of

CC the D-type cyclin cDNAs from a normal human liver genomic library.

CC A mutant yeast strain in which two of the three CLN cyclin genes

CC (CLN 1 and CLN 2) were inactivate and expression of the third was

CC conditional, was used to identify human cDNA clones that rescue yeast

CC from CLN deficiency. A human glioblastoma cDNA library carried in a

CC yeast expression vector (pADNS) was introduced into a mutant yeast

CC strain. Two yeast transformants (pCYCD1-21 and pCYCD1-19) which grew

CC despite the lack of function of all three CLN genes and were not

CC revertants, were identified and recovered in *E. coli*. These two

CC clones were shown to be independent clone representing the same gene.

CC A HeLa cDNA library was screened for a full-length cDNA clone using the

CC 1.2 kb insert of pCYCD1-21 as a probe. The sequence isolated by this

CC method was pCYCD1-H12 (see also Q31873). Degenerate probes and

CC primers were designed using the D1 gene sequence. These primers

CC and probes were used in the isolation of the cyclin D2 and D3 genes.

CC See also Q31874-5. The cyclin D1 cDNA clone was used to screen a

CC liver genomic library resulting in the identification of three

CC upstream promoters. These clones were shown to correspond to the

CC positive promoter region and a 198 bp exon, followed by an intron of

CC cyclin D1. Human cyclin promoters D2 and D3 were isolated in the same

CC manner. Cyclin D1 has been shown to be expressed differentially in

CC different cell types, with expression being highest in cells of neural

CC origin.

CC Sequence 3158 BP; 952 A; 674 C; 772 G; 910 A

```

Query Match          6.88; Score 135; DB 5; Length 3158;
Best Local Similarity 77.08; Pred. No. 1.64e-48;
Matches 221; Conservative 0; Mismatches 62; Indels 4; Gaps 2;

db 1653 gcccggaacggtggctcagcctgtatccacagcacttttgaggccgagacgcgcgga 1712
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
p 293 GGACGGGTATAGTGGCTACACCTATCTCCCAATGCTTTGGGAAGCCGAAGTGGGTGA 234
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

db 1713 tcac--gaggtcaggggttcaagactagcctggccaacatagtgaaaccccatctctacg 1770
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
p 233 TCACATGAGCCGAGGAGTGTGGAGACCGACCCCTGGCCACACATGGTGAACCCCTTATCTGCT 174
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db	1771	aaaaatacaaaaattagtcagcgcatggtgctgtcgttagctccagctactcggaa	1830
Cp	173	AAAAATACAAAATTAGTGGCATGTATACACACTGTATCCCACTATTGGGAA	114
Db	1831	ttgcttgaacccggaggttgaggttcagtgagccagatcgacacctgcactccagc	1890
Cp	113	TCACTTGAACCCAGAGGTGGAGTTTCAGTGAGCAAGATGCACACTGG -TCCAGC	56
Db	1891	ttgagcaacagtagactctgcctcaaaaaaaaaaaaaaaaaaaaaa	1937
Cp	55	CTGGGCACAGAGCAAGTCTCCCTCTCCGCCCAAAAAAATAAGAAA	9

RESULT 3
ID Q63862 standard; cDNA; 283 BP.
AC Q63862;
DT 29-JAN-1995 (first entry)
DE AP2 sequence obtd. by PCR for tumour specific DNA.
DE Arbitrary primers; AP-PCR; amplification; tumour cells; cancer;
KW Insertions; deletions; ss.
KW Synthetic.
OS WO9411531-A.
PN 26-MAY-1994.
PD 12-NOV-1993; U10904.
PF 12-NOV-1993; US-975737.
PR 13-NOV-1992; CALIFORNIA INST BIOLOGICAL RES.
PA (CALB-) Ikonov Y, Malkhosyan S, McClelland M, Peinado MA;
PI Peruchio M, Welsh;
PI WPI; 94-183529/22.
DR Identification of tumour cells - by analysing DNA to determine
PT whether insertions or deletions have occurred in reiterated
PT sequences
PT Disclosure: Page 52; 67pp; English.
PS The sequence was obtd. by PCR with arbitrary PCR primers used to
CC detect insertions or deletions in DNA sequences. Such mutations are
CC markers of cancer so such primers can be used in the diagnosis of
CC cancer, esp. colorectal, stomach or pancreatic tumours.
CC See also Q63837-63.
SQ sequence 283 BP: 63 A: 77 C: 94 C: 10 T: 1

Query Match	Score 121;	DB 11;	Length 283;
Best Local Similarity	84.68;	Pred. NO. 8,376-42;	
Matches 148;	Conservative 0;	Mismatches 27;	Indels 0; Gaps 0;
5	gggcgtggtgggtcacactgtaatccccagcacttttggagcgcgagtggtggatcac	64	
289	GGGTATAGTGCTCACACTATCTCCCAATGCTTTGGAAAGCCGAAGTGGGTGATCAC	230	
65	ctgaggtcaggaggttcaagaccagcctggccacaatggtgaaccccgctctactaaaa	124	
229	ATGAGCCAGGAGTTGGAGACCGCTGGCCACATGCTGAAACCCCTTATCTGCTAAA	170	
125	atacaaaaattagcgggcgtggttggcgcgcctgtaatcccgactactcggga	179	
169	ATACAAAAATTAGTGGGCGATGTTAAATACACACCTGTAAATCCCGAGCTATTTTGGGA	115	

RESULT	4	
D	Q94109	standard; DNA; 7849 BP.
D	Q94109;	
A	22-FEB-1996	(first entry)
T	DE	hML genomic DNA.
E	KW	Human; thrombopoietin; TPO; mpl ligand; hML; fragment polypeptide;
E	KW	megakaryocytopoietic cytokine receptor; thrombopoietic signal;
W	KW	EPO-domain fragment; erythropoietin; EPO; haematopoietic cell;
W	KW	megakaryocyte; thrombocytopenia; myeloproliferative disease;
W	KW	Inflammatory thrombocytosis; iron deficiency; EPO; platelet;
W	KW	red blood cell; progenitor; hML-2; ss.
S	H	Homo sapiens.
S	H	Location/Qualifiers
T	T	prim_transcript 1166..7289
T	T	/tag= a
T	T	exon 1161..1232
T	T	

MPSRELH

(TM)

Release 2.1D John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligentics, Inc.

MPSrch_n n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sun Jun 29 03:22:08 1997; Maspar time 1952.43 Seconds
Tabular output not generated. 1200.504 Million cell updates/sec

Title: >US-08-731-499-9
Description: (7201-9200) from US08731499.seq (5 of 6)
Perfect Score: 2006
N.A. Sequence: 7201 TATTAACCTCATATTCATCT.....AAGCAGAGCCCTCTCTGGCG 9200
Comp: ATATGAGGTATAAGTAGA.....TTCGTCTCGGAGGACCCG

Scoring table:

TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0.

Searched: 333249 seqs, 555961234 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

embi-new11
1: BCT 27: BCT2 27: BCT3 28: BCT4 29: BCT5 30: BCT6 31: BCT7
32: BCT8 33: BCT9 34: GEN1 35: GEN2 36: HTG 37: INV1 38: INV2
39: INV3 40: INV4 41: INV5 42: INV6 43: INV7 44: INV8 45: INV9
46: MAM1 47: MAM2 48: MAM3 49: VRT1 50: VRT2 51: VRT3 52: PAT1
53: PAT2 54: PAT3 55: PAT4 56: PHG 57: PLN1 58: PLN2 59: PLN3
60: PLN4 61: PLN5 62: PLN6 63: PLN7 64: PLN8 65: PLN9 66: PLN10
67: PRI1 68: PRI2 69: PRI3 70: PRI4 71: PRI5 72: PRI6 73: PRI7
74: PRI8 75: PRI9 76: PRI10 77: PRI11 78: PRI12 79: PRI13
80: PRI14 81: ROD1 82: ROD2 83: ROD3 84: ROD4 85: ROD5 86: ROD6
87: ROD7 88: ROD8 89: SVE 90: SYN 91: UNA 92: VRL1 93: VRL2
94: VRL3 95: VRL4 96: VRL5 97: VRL6 98: VRL7 99: VRL8 100: VRL9

Database:

genbank-new11
101: BCT 102: GEN 103: INV1 104: INV2 105: MAM 106: VRT
107: PHG 108: PLN 109: PRI1 110: PRI2 111: ROD 112: SYN
113: UNA 114: VRL
u-emb148_97
115: part1 116: part2

Database:

Statistics: Mean 12.230; Variance 6.274; scale 1.949

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
Query						

1	204	10.2	152141	73	HSU07000	Human breakpoint clus	2.15e-130
2	202	10.1	2928	70	HSAP0A2G	Human gene for apolip	8.49e-129
3	203	10.1	3360	70	HSAP0A2G	Human gene for apolip	8.49e-129
4	203	10.1	158827	110	HSK217C2	Human DNA sequence fr	1.35e-129
5	200	10.0	7407	76	HUMGIPRD	Human gene for gastr	3.34e-127
6	200	10.0	26764	76	HUMDODA	Homo sapiens deoxycyt	3.34e-127
7	200	10.0	37194	75	HSV210E9	Human DNA sequence fr	3.34e-127
8	200	10.0	84539	70	HSABLGR3	Human proto-oncogene	3.34e-127
9	200	10.0	116879	74	HSU47924	Human chromosome 12p1	3.34e-127
10	200	10.0	222930	109	HSU47924	Human chromosome 12p1	3.34e-127
11	197	9.9	4208	69	HS91K3D	Human DNA sequence fr	8.20e-125
12	197	9.9	36594	79	HUMTPA	Human tissue plasmino	8.20e-125
13	197	9.9	38538	70	HSN95B1	Human DNA sequence fr	8.20e-125
14	198	9.9	40227	72	HSN119A7	Human DNA sequence fr	1.31e-125
15	198	9.9	40227	110	HSN119A7	Human DNA sequence fr	1.31e-125
16	198	9.9	40509	75	HSU73E8	Human DNA sequence fr	1.31e-125
17	197	9.9	49625	6	HSB13C9	Human DNA sequence fr	1.31e-125
18	197	9.9	49625	4	HSB13C9	Human DNA sequence fr	1.31e-125
19	197	9.9	49625	110	HSB13C9	Human DNA sequence fr	8.20e-125
20	197	9.9	49625	5	HSB13C9	Human DNA sequence fr	8.20e-125
21	198	9.9	90000	36	HSAC000010	EPM1/APECED region of	1.31e-125
22	198	9.9	90000	4	HSAC10	EPM1/APECED region of	1.31e-125
23	198	9.9	90000	6	HSAC10	EPM1/APECED region of	1.31e-125
24	198	9.9	90000	5	HSAC10	EPM1/APECED region of	1.31e-125
25	199	9.9	115062	110	HSK799F10	Human DNA sequence fr	2.09e-126
26	197	9.9	120000	5	HSAC00003	Chromosome 17 genomic	8.20e-125
27	197	9.9	120000	36	HSAC000003	Chromosome 17 genomic	8.20e-125
28	197	9.9	120000	4	HSAC00003	Chromosome 17 genomic	8.20e-125
29	197	9.9	120000	6	HSAC00003	Chromosome 17 genomic	8.20e-125
30	198	9.9	140787	110	HSJ293L6	Human DNA sequence fr	1.31e-125
31	199	9.9	144401	110	HSJ104C13	Human DNA sequence fr	1.31e-125
32	199	9.9	155821	110	HSJ78B3	Human DNA sequence fr	2.09e-126
33	199	9.9	164393	110	HSJ272J12	Human DNA sequence fr	2.09e-126
34	196	9.8	16397	71	HSJ60G9B	Human DNA sequence fr	5.13e-124
35	196	9.8	30110	71	HSJ124H12	Human DNA sequence fr	5.13e-124
36	196	9.8	37194	75	HSV210E9	Human DNA sequence fr	5.13e-124
37	195	9.8	40176	109	HSU58675	Human olfactory recep	3.20e-123
38	196	9.8	40905	110	HSN87F1	Human DNA sequence fr	5.13e-124
39	195	9.8	44888	67	HS130N4B	Human DNA sequence fr	3.20e-123
40	195	9.8	44888	78	HOMMMDBC	Human DNA from cosmid	3.20e-123
41	196	9.8	95819	110	HSK286B10	Human DNA sequence fr	5.13e-124
42	195	9.8	106000	74	HSU40455	Human chromosome X co	3.20e-123
43	195	9.8	111344	4	HS17	Human DNA sequence fr	3.20e-123
44	195	9.8	111344	6	HS17	Human DNA sequence fr	3.20e-123
45	196	9.8	163037	110	HSJ6802	Human DNA sequence fr	5.13e-124

ALIGNMENTS

RESULT 1

LOCUS

DEFINITION

ACCESSION

NID

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

HSU07000 152141 bp DNA PRI 17-JAN-1996
Human breakpoint cluster region (BCR) gene, complete cds.
007000
9487344

human.

Homo sapiens

Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 152141)

Chisoe, S.L., Bodenteich, A., Wang, Y.F., Wang, Y.P., Burlan, D.,

Clifton, S.W., Crabtree, J., Freeman, A., Iyer, K., Jian, L., Ma, Y.,

McLaurin, H.-J., Pan, H.-Q., Sarhan, O.H., Toth, S., Wang, Z., Zhang, G.,

Heisterkamp, N., Groffen, J., and Roe, B.A.

Sequence and analysis of the human ABL gene, the BCR gene, and

regions involved in the Philadelphia chromosomal translocation

Genomics 27 (1), 67-82 (1995)

9539474

2 (bases 14590 to 16317; 87877 to 88058; 95028 to 95132; 95433 to

95618; 102486 to 102593; 105610 to 105670; 107159 to 107211; 107712

to 107852; 118055 to 118176; 119111 to 119279; 121237 to 121356;

122175 to 122250; 123595 to 123699; 124417 to 124491; 126619 to

JOURNAL MEDLINE REFERENCE AUTHORS	library 216 (4550), 1136-1138 (1982) Science 8219944 24 (sites) Soekarman, D., van Denderen, J., Hoefsloot, L., Moret, M., Meeuwse, T., van Baal, J., Hagemeijer, A. and Grosveld, G. A novel variant of the bcr-abl fusion product in Philadelphia chromosome-positive acute lymphoblastic leukemia Leukemia 4 (6), 397-403 (1990) 90294679 25 (sites) Steiterkamp, N., Stephenson, J. R., Groffen, J., Hansen, P. F., de Klein, A., Bartam, C. R. and Grosveld, G. Localization of the c-abl oncogene adjacent to a translocation break point in chronic myelocytic leukaemia Nature 306 (5940), 239-242 (1983) 84068136 25 (sites) Grosveld, G., Verwoerd, T., van Agthoven, T., de Klein, A., Ramachandran, K. B., Heisterkamp, N., Stam, K. and Groffen, J. The chronic myelocytic cell line K562 contains a breakpoint in bcr and produces a chimeric bcr/c-abl transcript Mol. Cell. Biol. 6 (2), 607-616 (1986) 87064346 27 (sites) Morris, C. M., Heisterkamp, N., Groffen, J. and Fitzgerald, P. H. Entire ABL gene is joined with 5'-BCR in some patients with Philadelphia-positive leukemia <u>Blood 78 (4), 1078-1084 (1991)</u> 91329820 JOURNAL MEDLINE REFERENCE AUTHORS	Heisterkamp, N., Knoppel, E. and Groffen, J.
--	--	--

```

...: remainder of annotations omitted.
Query Match      10-24: Score 204; DB 73; Length 152141;
Best Local Similarity (88.7%);
Matches      251; Conservativeness 0; Mismatches 29; Indels 3; Gaps 3;

Db 86498 ttcttttctttctttttttt-ttttttgagcagagctctgtgctgtgcacctaggct 86556
QY 7427 TTCATTCTTTCTTTCTTTTGTGTTTTTGAGACGGAGCTTGTCTGTCACTCGCT 7486

Db 86557 ggagtcagtggtgcaatctgtctcactcaacctccactccacaggttcaagcgattc 86616
QY 7487 GGAGTGCAGTGGTGCATATTTGGCTCTACTGCAGCCCTCACCTCCTGGTTCAAGCAATAC 7546

Db 86617 tctgtctcagctcctcgtagtagctgggattacaggcaactgcacacatgcccacctaat 86676
QY 7547 TCCTGGCTCAGCTCCCCAGTGTGGGATTACAGGTACCTGCCACCACACCCCGGCTAAT 7606

Db 86677 tttttatagtttagtagacatggggtt-caccatatggccagcgtggtctcgaaatcc 86735
QY 7607 TTTCTG-TATTTTATAGTAGAATGGGGTTTCACCATGTGGCCAGGCTGGTTTCGAACATCC 7665

Db 86736 tgacctcaagtgcaccaccaacctcggcctcccaagtgcctgg 86778
QY 7666 TGACCTCAAGTGATCCGCTCACCTTGGCTCCCATAGTGTGG 7708

```

2
RESULT HSAPOA2G 12928 bp DNA
LOCUS Human gene for apolipoprotein AII.
DEFINITION
ACCESSION X02905
NID 928757
KEYWORDS AII repetitive sequence; apolipoprotein A-II; inverted repeat;
terminal repeat; 2. DNA.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2928)
AUTHORS Knott,T.J., Wallis,S.C., Robertson,M.E., Priestley,L.M., Urdea,M.

TITLE		The human apolipoprotein AII gene: structural organization
JOURNAL MEDLINE	86016095	Nucleic Acids Res. 13 (17), 6387-6398 (1985)
REFERENCE AUTHORS TITLE	2 (bases 715 to 2456) Lackner,K.J., Law,S.W. and Brewer,H.B. Jr. The human apolipoprotein A-II gene: complete nucleic acid and genomic organization	(1985)
JOURNAL MEDLINE	85242123	Nucleic Acids Res. 13 (12), 4597-4608 (1985)
COMMENT FEATURES	Data kindly reviewed	(18-DEC-1985) by J. Scott.
source	Location/Qualifiers	
repeat_region	1..2928	/organism="Homo sapiens"
	556..565	/note="direct repeat 1"
repeat_region	573..582	/note="direct repeat 2"
repeat_region	591..600	/note="direct repeat 2"
repeat_region	607..616	/note="direct repeat 1"
repeat_region	703..712	/note="direct repeat 3"
repeat_region	730..739	/note="direct repeat 3"
CAAT_signal	834..842	/note="put. CAAT-like sequence"
TATA_signal	883..888	/number=1
exon	912..945	join(912..945,1115..1190,1484..1616,2012..2241)
mRNA	prim.transcript	912..2241
misc_RNA	912	/note="major cap site"
misc_RNA	916	/note="minor cap site"
conflict		replace(943..944,"tagtaagataagga") /note="ta is taggtaagataagga in [2]" citation=[2] 946..1114
intron		/note="intron I" replace(989,"t") /note="c is t in [2]" citation=[2]
conflict		replace(1019,"t") /note="c is t in [2]" citation=[2] 1115..1190
exon		/number=2
sig_peptide		join(1139..1190,1484..1485)
CDS		join(1139..1190,1484..1616,2012..2129) /codon_start=1 /product="apo AII" /db_xref="PID:g296634" /db_xref="SWISS-PROT:P02652" /translation="MKLIATYVLLITCSLEGALVRQAKEPCVESI" GKDLMEKYGPQLQAERAKSFKEKSQELPLIKAGTELVNFLSYFV" 1191..1483
intron		/note="intron II"
conflict		replace(1415,"g") /note="a is g in [2]" citation=[2] 1446..1477
misc_feature		/note="(GT)x16, pot. zDNA structure"
exon		1484..1616
mat_peptide		/number=3 join(1501..1616,2012..2126) /product="mature apo AII" replace(1530,"g") /note="a is g in [2]" citation={2}
conflict		